ECONSTOR

WWW.ECONSTOR.EU

Der Open-Access-Publikationsserver der ZBW – Leibniz-Informationszentrum Wirtschaft
The Open Access Publication Server of the ZBW – Leibniz Information Centre for Economics

Bissantz, Nicolai; Dümbgen, Lutz; Holzmann, Hajo; Munk, Axel

Working Paper

Nonparametric confidence bands in deconvolution density estimation

Technical Report // Sonderforschungsbereich 475, Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, No. 2007,03

Provided in cooperation with:
Technische Universität Dortmund

Suggested citation: Bissantz, Nicolai; Dümbgen, Lutz; Holzmann, Hajo; Munk, Axel (2007) : Nonparametric confidence bands in deconvolution density estimation, Technical Report // Sonderforschungsbereich 475, Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, No. 2007,03, http://hdl.handle.net/10419/36581

Nutzungsbedingungen:
Die ZBW räumt Ihnen als Nutzerin/Nutzer das unentgeltliche, räumlich unbeschränkte und zeitlich auf die Dauer des Schutzrechts beschränkte einfache Recht ein, das ausgewählte Werk im Rahmen der unter
→ http://www.econstor.eu/dspace/Nutzungsbedingungen
nachzulesenden vollständigen Nutzungsbedingungen zu vervielfältigen, mit denen die Nutzerin/der Nutzer sich durch die erste Nutzung einverstanden erklärt.

Terms of use:
The ZBW grants you, the user, the non-exclusive right to use the selected work free of charge, territorially unrestricted and within the time limit of the term of the property rights according to the terms specified at
→ http://www.econstor.eu/dspace/Nutzungsbedingungen
By the first use of the selected work the user agrees and declares to comply with these terms of use.

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Mitglied der Leibniz-Gemeinschaft

Nonparametric confidence bands in deconvolution density estimation

Nicolai Bissantz[1,2], Lutz Dümbgen[3], Hajo Holzmann[1] and Axel Munk[1]

[1] Institute for Mathematical Stochastics

Georg-August-University at Göttingen

Germany

[2] Faculty of Mathematics

Ruhr-University Bochum

Germany

[3] Institute for Mathematical Statistics and Actuarial Science

University of Bern

Switzerland

Abstract

Uniform confidence bands for densities f via nonparametric kernel estimates were first constructed by Bickel and Rosenblatt [*Ann. Statist.* **1**, 1071–1095]. In this paper this is extended to confidence bands in the deconvolution problem $g = f * \psi$ for an ordinary smooth error density ψ. Under certain regularity conditions, we obtain asymptotic uniform confidence bands based on the asymptotic distribution of the maximal deviation (L_∞-distance) between a deconvolution kernel estimator $\hat{f}$ and f. Further consistency of the simple nonparametric bootstrap is proved. For our theoretical developments the bias is simply corrected by choosing an undersmoothing bandwidth. For practical purposes we propose a new data-driven bandwidth selector based on heuristic arguments, which aims

[1]Adress for correspondence: Dr. Nicolai Bissantz, Faculty of Mathematics, Ruhr-University Bochum, Universitätsstraße 150, Mathematik III NA 3/70, D-44780 Bochum, Germany, email: nicolai.bissantz@rub.de, Fon: +49/234/32–14559, Fax: +49/234/32–23291

at minimizing the L_∞-distance between $\hat{f}$ and f. Although not constructed explicitly to undersmooth the estimator, a simulation study reveals that the suggested bandwidth selector performs well in finite samples, both in terms of area and coverage probability of the resulting confidence bands. Finally the methodology is applied to measurements of the metallicity of local F and G dwarf stars. Our results confirm the "G dwarf problem", i.e. the lack of metal-poor G dwarfs relative to predictions from "closed-box models" of stellar formation.

Keywords: astrophysics; asymptotic normality; bootstrap; confidence band; deconvolution; nonparametric density estimation; bandwidth selection

AMS subject classification: Primary 62G15; Secondary 62G07.

1 Introduction

Nonparametric function estimation is an important tool for analyzing data, both for purposes of statistical inference as well as for graphical visualization. In the latter context, one has to distinguish whether a feature of the curve estimate is only due to random fluctuations, or whether it captures relevant structure of the unknown curve. To this end, interval estimates are frequently employed, and uniform confidence intervals (i.e. confidence bands) seem to be more appropriate than pointwise confidence intervals.

In a pioneering work, Bickel and Rosenblatt (1973) constructed confidence bands for a density function of i.i.d. observations, based on the asymptotic distribution of the supremum of a centered kernel density estimator. Since then, their method has been further developed both in the density estimation and also in a regression framework. For example, Giné, Koltchinskii and Sakhanenko (2004) derived the asymptotic distributions of weighted supremum-type statistics for kernel density estimators. Hall (1993) investigated bootstrap confidence bands, and Hall and Owen (1993) used empirical likelihood methods. In a regression context, Härdle (1989) constructed asymptotic confidence bands for M-smoothers. Eubank and Speckman

(1993) for the Nadaraya-Watson estimator and Xia (1998) for local polynomial estimators, respectively, suggested confidence bands based on an explicit bias correction and not on undersmoothing. Bootstrap confidence bands for nonparametric regression were proposed by Neumann and Polzehl (1998) and by Claeskens and van Keilegom (2003).

Although there is quite a variety of methods for constructing confidence bands in direct regression and density estimation models, in indirect models such as inverse regression or deconvolution density estimation there seem to be no techniques available yet. Our goal in this paper is to partially fill this gap by constructing confidence bands in ordinary smooth deconvolution problems.

To be more precise, let $X_1, \ldots, X_n$ be i.i.d. observations from the model

$$X_i = Z_i + \epsilon_i, \tag{1}$$

where we assume that the ϵ_i are i.i.d. with known density ψ and independent of the Z_i. The object of interest is the density f of the Z_i, which is related to the density g of the X_i via

$$g = f * \psi, \tag{2}$$

the convolution of f and ψ. Recovering f from the noisy observations $X_1, \ldots, X_n$ is therefore called the deconvolution problem (e.g. Fan, 1991a/b; Diggle & Hall, 1993; Delaigle & Gijbels, 2002). It is well-known that the optimal rate at which f can be estimated depends on the smoothness of f as well as on the smoothness of the error density ψ. To fix the notation, denote the Fourier transform of f by $\Phi_f(t) = \int_{\mathbb{R}} f(x) \exp(itx)\, dx$. Roughly speaking, the error density is called ordinary smooth if its Fourier transform $|\Phi_\psi(t)|$ decays at a polynomial rate as $t \to \infty$, in which case the optimal rate for estimating f is also of polynomial order. In contrast, if $|\Phi_\psi(t)|$ decays at an exponential rate as $t \to \infty$, ψ is called supersmooth and the optimal rate for f is typically only of logarithmic order (an exception is the case in which f is also supersmooth). For details see Fan (1991a) and Pensky and Vidakovic (1999), among others.

In the deconvolution problem, several nonparametric estimators for f are available: kernel-based estimators (e.g. Stefanski and Carroll, 1990), estimators based on wavelets (Pensky and Vidakovic, 1999) or iterative methods (Hesse and Meister, 2004). Here we will restrict

ourselves to kernel estimators. Suppose that f is p-times continuously differentiable for some $p \geq 0$. Under the assumption that $\Phi_\psi(t) \neq 0$ for all $t \in \mathbb{R}$ and that Φ_k, the Fourier transform of the kernel $\mathcal{K}$, has compact support, the kernel deconvolution density estimator for the j^{th} derivative of f, given by,

$$\hat{f}_n^{(j)}(x) = \hat{f}_{n,h}^{(j)}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (-it)^j \, \exp(-itx) \, \Phi_k(ht) \frac{\hat{\Phi}_n(t)}{\Phi_\psi(t)} \, dt, \qquad 0 \leq j \leq p, \tag{3}$$

is well-defined. Here $h > 0$ is a smoothing parameter called bandwidth, and $\hat{\Phi}_n$ is the empirical characteristic function of $X_1, \ldots, X_n$.

The estimator $\hat{f}_n^{(j)}$ can be written in kernel form as follows:

$$\hat{f}_n^{(j)}(x) = \frac{1}{nh^{j+1}} \sum_{k=1}^{n} K_n^{(j)}\Big(\frac{x - X_k}{h}\Big), \tag{4}$$

where

$$K_n^{(j)}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} (-it)^j \, \exp(-itx) \frac{\Phi_k(t)}{\Phi_\psi(t/h)} \, dt \quad 0 \leq j \leq p. \tag{5}$$

Notice that $K_n^{(j)}$ depends on n through h. Asymptotic normality of $\hat{f}_n^{(j)}(x)$ was derived by Fan (1991b) and van Es and Uh (2005), both in the ordinary smooth and in the super smooth cases under some regularity conditions. Practical suggestions of how to choose the bandwidth in order to minimize the MISE are given in Delaigle and Gijbels (2004).

In this paper we construct confidence bands for $f^{(j)}$ in the ordinary smooth deconvolution problem. In Section 2 we discuss regularity properties related to an ordinary smooth error density ψ. In particular, we will require that the Fourier transform of ψ decays exactly at a polynomial rate. Such a condition was also used by Fan (1991b) to prove asymptotic normality of $\hat{f}_n^{(j)}(x)$. In Section 3 we obtain asymptotic confidence bands by using the method of Bickel and Rosenblatt (1973). Our regularity condition on the error density ψ guarantees that the deconvolution kernel (5) has a simple asymptotic expression (cf. equation (8)), which makes the Bickel-Rosenblatt argument applicable. In Section 4 we construct bootstrap confidence bands via the simple non-parametric bootstrap. Hall (1993) showed in the direct density estimation context that bootstrap confidence bands have a much better coverage accuracy than asymptotic confidence bands. His arguments are based on the Edgeworth expansion. However, in the deconvolution problem such elaborate arguments seem to be rather inaccessible.

Therefore we give a simpler and direct proof of consistency of the bootstrap which relies on a method to show consistency of the bootstrap of the empirical process due to Shorack (1982). In our theoretical developments we simply correct the bias by choosing an undersmoothing bandwidth.

Section 5 contains an extensive simulation study. After an introductory outline of the simulation framework in Section 5.1, for practical purposes we introduce in Section 5.2 a new data-driven bandwidth selection procedure. Heuristic arguments indicate that the procedures aims at minimizing the L_∞-distance between the density estimator and the unknown true density f. This is in accordance with our aim to construct a confidence band in the L_∞-norm. While we do not provide a theoretical justification for this bandwidth selector and in particular, do not construct it explicitly to undersmooth the estimator, the simulation study reveals that the resulting confidence bands perform well in terms of coverage probability and area. Let us mention that most bandwidth selection methods suggested in the literature for deconvolution problems have a tendency to oversmooth the estimator. While these bandwidths lead to satisfactory results for estimating f itself, the resulting confidence bands typically have coverage probabilities far below their nominal levels. In Section 5.3, we compare the performance of asymptotic confidence bands and of confidence bands based on the bootstrap. Finally, in Section 5.4 we investigate the sensitivity of the constructed confidence bands w.r.t. misspecification of the error distribution.

This work is motivated by an astrophysical problem, namely the estimation of the density of metallicities of F and G dwarfs, i.e. of stars very similar to the Sun. In Section 6 we compute bootstrap confidence bands for estimates of this density for a sample of such stars in the Solar neighbourhood, and confirm the "G-dwarf" problem (see e.g. van den Bergh, 1962; Schmidt, 1963; Sommer-Larsen, 1961), which is characterized by a substantial lack of low metallicity F- and G-dwarf stars as compared to certain standard theories of star formation. It is important to note that our analysis confirms this conclusion for the data after correction (deconvolution) for measurement errors in the observed metallicities. Moreover, observing the confidence bands for the estimator of the metallicity density it appears to be very unlikely that the lack of metal-poor G dwarfs is due to observational errors.

Finally, all proofs are deferred to an appendix.

2 The ordinary smooth deconvolution problem

In this section we discuss regularity properties of the deconvolution problem (2) in the ordinary smooth case, that will be of importance subsequently, and also introduce some relevant notation. In general, the deconvolution problem is called ordinary smooth if the characteristic function Φ_ψ of the error variable ϵ satisfies $\Phi_\psi(t) \neq 0$ for all $t \in \mathbb{R}$, and if there exist $c, C, \beta > 0$ such that

$$c|t|^{-\beta} \leq |\Phi_\psi(t)| \leq C|t|^{-\beta}, \tag{6}$$

for $|t|$ sufficiently large. Minimax-rates in the deconvolution problem can be derived under (6), c.f. Fan (1991a) or Mair and Ruymgaart (1996). However, to obtain an explicit asymptotic formula for the variance of the estimator (3) as well as its asymptotic normality, the following stronger requirement becomes necessary,

$$\Phi_\psi(t)t^\beta \to C_\epsilon, \qquad t \to \infty, \tag{7}$$

for some $\beta \geq 0$ and $C_\epsilon \in \mathbb{C} \setminus \{0\}$. Note that this implies that $\Phi_\psi(t)|t|^\beta \to \bar{C}_\epsilon$, $t \to -\infty$. If (7) holds, the deconvolution kernel $K_n^{(j)}$ given in (5) has a simple asymptotic form. In fact, from the dominated convergence theorem,

$$h^\beta K_n^{(j)}(x) \to K^{(j)}(x), \quad h \to 0,$$

where

$$\begin{aligned} K^{(j)}(u) &= \frac{1}{2\pi C_\epsilon} \int_0^\infty (-ix)^j \exp(-iux) x^\beta \Phi_k(x)\, dx \\ &\quad + \frac{1}{2\pi \overline{C_\epsilon}} \int_{-\infty}^0 (-ix)^j \exp(-iux)|x|^\beta \Phi_k(x)\, dx, \end{aligned} \tag{8}$$

c.f. Fan and Liu (1998). Note that the second term in (8) is the complex conjugate of the first, so that $K^{(j)}(u)$ is in fact real-valued. This allows e.g. to obtain an explicit asymptotic formula for the pointwise variance of the estimator (3), which is proportional to $h^{-2\beta-2j-1}\, n^{-1}\, g(x)$. No such formula seems to be available in the general case (6), and in fact the order of the

Table 1

	density	characteristic fct.	β	C_ϵ
Gamma distr.	$x^{\alpha-1}\exp(-x)/\Gamma(\alpha),\ x,\alpha>0$	$(1-it)^{-\alpha}$	α	i^α
Laplace distr.	$\exp(-\lvert x\rvert)/2$	$1/(1+t^2)$	2	1

Table 1: Error densities, characteristic functions and constants

variance is not clear. Therefore, in order to construct confidence bands, we will require the stronger regularity assumption (7).

Let us mention that for certain supersmooth deconvolution problems, in particular for normal deconvolution, there is also an asymptotic formula for the pointwise variance (cf. van Es and Uh, 2005) available. However, e.g. in case of normal deconvolution, this limit variance no longer depends on f and x, but only (in a global way) on the error distribution. Furthermore, there is no analogue of the limit form (8) of the deconvolution kernel $K_n^{(j)}$ available. Therefore, our methods do not extend to this case, and in fact we conjecture that the asymptotic distributions of supremum-type statistics might be different.

To conclude this section let us consider two examples.

Example 1. Suppose that Φ_ψ is the reciprocal of a polynomial, i.e.

$$\frac{1}{\Phi_\psi(t)} = \sum_{j=0}^{\beta} a_j t^j, \quad a_j \in \mathbb{C}.$$

It is evident that (7) is satisfied with $C_\epsilon = 1/a_\beta$. Specific examples are given in Table 1.

Example 2. Let ψ_0 be a density with characteristic function satisfying (7), and consider the mixture

$$\psi(x) = \lambda\psi_0(x-\mu)/2 + \lambda\psi_0(x+\mu)/2 + (1-\lambda)\psi_0(x)$$

for some $0<\lambda<1/2$ and $\mu\neq 0$. In this case the resulting characteristic function is

$$\Phi_\psi(t) = \big(1-\lambda+\lambda\cos(\mu t)\big)\Phi_{\psi_0}(t).$$

Here (6) is satisfied, but (7) is not.

3 Asymptotic confidence bands

In this section we construct asymptotic confidence bands on a compact interval in the ordinary smooth case. For simplicity we formulate our results for the interval $[0,1]$, however, they can be easily extended to any compact interval $[a,b] \subset \mathbb{R}$. We start by studying the distribution of the maximum of the process

$$Y_n^{(j)}(t) = \frac{n^{1/2} h^{\beta+j+1/2}}{g(t)^{1/2}} \big(\hat{f}_n^{(j)}(t) - E\hat{f}_n^{(j)}(t) \big), \quad t \in [0,1].$$

Let us list the regularity conditions that we require in the following.

Assumption 1. The Fourier transform Φ_k of k is symmetric, three times differentiable and supported on $[-1,1]$, $\Phi_k(t) = 1$ for $t \in [-c,c]$, $c > 0$, and $|\Phi_k(t)| \le 1$.

Assumption 2. A. $\int_{\mathbb{R}} |K_n^{(j+1)}(x)||x|^{3/2} (\log\log^+ |x|)^{1/2}\, dx = O(h^{-\beta})$, where $\log\log^+ |x| = 0$ if $|x| < e$, and $\log\log^+ |x| = \log\log |x|$, otherwise.

B. For some $\delta > 0$,

$$\int_{\mathbb{R}} |h^\beta K_n^{(j+1)}(x) - K^{(j+1)}(x)||x|^{1/2} (\log\log^+ |x|)^{1/2}\, dx = O(h^{1/2+\delta}),$$

where $K^{(j+1)}$ is given in (8).

Assumption 3. A. The density g is bounded and bounded away from 0 on $[0,1]$. Furthermore, $g^{1/2}$ is differentiable with bounded derivative.

B. The Fourier transform Φ_f of f satisfies

$$\int_{\mathbb{R}} |\Phi_f(t)||t|^{r-1}\, dt < \infty \qquad \text{for some} \quad r > p+1.$$

We will discuss these assumptions and how they can be verified subsequently in Remark 1. Now we are in the position to state the following limit theorem. Let $\|\cdot\|_I$ denote the sup-norm on an interval $I \subset \mathbb{R}$.

Theorem 1. *Under the Assumptions 1 – 3, if $h \to 0$ and $hn/\log^3(n) \to \infty$, then for $0 \le j \le p$,*

$$P\Big(\big(2\log(1/h)\big)^{1/2} \big(\|Y_n^{(j)}\|_{[0,1]} / C_{K,1}^{1/2} - d_n \big) \le x \Big) \to \exp(-2\exp(-x)),$$

where

$$d_n = \big(2\log(1/h)\big)^{1/2} + \frac{\log\big(\frac{1}{2\pi}C_{K,2}^{1/2}\big)}{\big(2\log(1/h)\big)^{1/2}},$$

and

$$C_{K,1} = \frac{1}{2\pi|C_\epsilon|^2}\int_{\mathbb{R}} x^{2(\beta+j)}\Phi_k^2(x)\,dx, \quad C_{K,2} = \frac{\int_{\mathbb{R}} x^{2(\beta+j+1)}\Phi_k^2(x)\,dx}{\int_{\mathbb{R}} x^{2(\beta+j)}\Phi_k^2(x)\,dx}. \tag{9}$$

From Theorem 1 one can directly construct confidence bands for the smoothed version $E\hat{f}_n^{(j)}$ of $f^{(j)}$. In order to construct confidence bands for $f^{(j)}$ itself, we have to deal with the bias of $\hat{f}_n^{(j)}$. Now $E\hat{f}_n^{(j)}(x) = k_h * f^{(j)}(x)$, $k_h(x) = k(x/h)/h$, does not depend on the error distribution, and in principle we could expand the bias and use an explicit bias correction by estimating its dominating term (see Eubank and Speckman, 1993; Xia, 1998, for bias-corrected confidence bands in a regression context). However, this strategy has specific disadvantages in the deconvolution context, for a discussion see Remark 1 and Section 5. Therefore, for our theoretical developments we correct the bias by choosing an undersmoothing bandwidth. Using Assumptions 1 and 3 B., one can derive the following estimate for the bias

$$\big|E\hat{f}_n^{(j)}(x) - f^{(j)}(x)\big| = \frac{1}{2\pi}\Big|\int_{\mathbb{R}}(-it)^j\,\exp(-itx)(1-\Phi_k(ht))\Phi_f(t)\,dt\Big| = o(|h|^{r-j-1}), \tag{10}$$

uniformly in x. Moreover we have to estimate the unknown density g of the observations. We require an estimator $\tilde{g}_n$ of g which satisfies

$$\|\tilde{g}_n(t) - g(t)\|_{[0,1]} = o_P\big((\log(1/h))^{-1}\big), \tag{11}$$

where h is the bandwidth used to estimate f. In this way we get the following asymptotic confidence bands.

Corollary 2. *Let $\tilde{g}_n$ be an estimator of g satisfying (11). Under the Assumptions 1 – 3, if $nh^{2(\beta+j)+1}/\log(1/h) \to \infty$ and $nh^{2\beta+2r-1}\,\log(1/h) \to 0$, we have*

$$P\big(\hat{f}_n^{(j)}(t) - b_n(t,x) \le f^{(j)}(t) \le \hat{f}_n^{(j)}(t) + b_n(t,x) \text{ for all } t \in [0,1]\big) \to \exp(-2\exp(-x)), \tag{12}$$

where

$$b_n(t,x) = \Big(\frac{\tilde{g}_n(t)C_{K,1}}{nh^{2(\beta+j)+1}}\Big)^{1/2}\Big(\frac{x}{(2\log(1/h))^{1/2}} + d_n\Big).$$

Notice that the width of the bands is of order $\big(\log(1/h)/(nh^{2\beta+2j+1})\big)^{1/2}$, thus the first condition on the bandwidth ensures that this converges to 0. Furthermore, it implies that the bandwidth assumption $hn/\log n \to \infty$ is satisfied. We notice that the slower $h \to 0$, the smaller the band. However, since we undersmooth in order to correct for the bias, h has to tend to 0 in such a way that $nh^{2\beta+2r-1}\log(1/h) \to 0$. Notice that both conditions can be met simultaneously since $r > p+1$ and $j \leq p$.

Remark 1 (*Discussion of the assumptions).* Here we discuss the Assumptions 1 – 3, and how they can be verified.

Kernels satisfying Assumption 1 are called flat-top kernels (cf. Politis and Romano, 1999). They have good bias properties in the sense that in contrast to kernels of finite order, they achieve optimal rates for all possible degrees of smoothness $r > p+1$ of the density as shown in (10). However, as also indicated in (10), for flat-top kernels there is no simple form for the leading term of the bias, and therefore explicit bias correction is not possible. For ordinary smooth deconvolution as considered in this paper, one could also simply use the Gaussian kernel. If f is at least $j+2$ times differentiable, then for the Gaussian $E\hat{f}_n^{(j)}(x) = k_h * f^{(j)}(x)$ is of order h^2, and its leading term could be estimated by estimating $f^{(j+2)}$. However, in a simulation study it turned out that when using the Gaussian kernel there is not practical gain in explicit bias correction for deconvolution problems, since estimating $f^{(j+2)}$ is an even more sensitive matter than estimating $f^{(j)}$.

Next let us consider Assumption 2. As for Assumption 2 A., if Φ_ψ is three times continuously differentiable, from twofold partial integration we obtain

$$\begin{aligned} K_n^{(j+1)}(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} (-it)^{j+1}\exp(-itx)\frac{\Phi_k(t)}{\Phi_\psi(t/h)}\,dt \\ &= \frac{(-i)^{j-1}}{2\pi x^2}\int_{\mathbb{R}} \exp(-itx)\left(\frac{t^{j+1}\Phi_k(t)}{\Phi_\psi(t/h)}\right)''\,dt. \end{aligned}$$

Therefore it suffices to show that

$$\int_{\mathbb{R}}\Big|\frac{t^{j+1}\Phi_k(t)}{\Phi_\psi(t/h)}\Big|dt = O\big(h^{-\beta}\big) \quad \text{and} \quad \int_{\mathbb{R}}\Big|\left(\frac{t^{j+1}\Phi_k(t)}{\Phi_\psi(t/h)}\right)''\Big|\,dt = O\big(h^{-\beta}\big). \tag{13}$$

These properties will typically be satisfied for a general ordinary smooth deconvolution prob-

lem, i.e. if (6) holds. Concerning Assumption 2 B., we have

$$|h^\beta K_n^{j+1}(x) - K^{j+1}(x)| \leq \frac{1}{2\pi}\Big(\int_0^\infty |t^{j+1}\Phi_k(t)| \Big| \frac{h^\beta}{\Phi_\psi(t/h)} - \frac{t^\beta}{C_\epsilon} \Big| \, dt + \int_{-\infty}^0 |t^{j+1}\Phi_k(t)| \Big| \frac{h^\beta}{\Phi_\psi(t/h)} - \frac{|t|^\beta}{C_\epsilon} \Big| \, dt \Big) \tag{14}$$

and

$$|h^\beta K_n^{j+1}(x) - K^{j+1}(x)| \leq \frac{1}{2\pi x^2}\Big(\int_0^\infty \Big| \Big(t^{j+1}\Phi_k(t)\big(\frac{h^\beta}{\Phi_\psi(t/h)} - \frac{t^\beta}{C_\epsilon}\big)\Big)^{(2)} \Big| \, dt + \int_{-\infty}^0 \Big| \Big(t^{j+1}\Phi_k(t)\big(\frac{h^\beta}{\Phi_\psi(t/h)} - \frac{|t|^\beta}{C_\epsilon}\big)\Big)^{(2)} \Big| \, dt \Big). \tag{15}$$

Therefore we only have to check that the integrals in (14) and (15) are of order $O(h^{1/2+\delta})$. Note that regularity property (7) together with the dominated convergence theorem implies that (14) tends to zero. In order to additionally obtain the rate $h^{1/2+\delta}$ in (14), one needs a stronger version of (7), e.g. $\big(\Phi_\psi(t)t^\beta - C_\epsilon\big)t^{1/2+\beta} \to 0$, $t \to \infty$. Let us mention that Assumption 2 B. in fact implies (7). Therefore, Assumption 2 B. can be considered as a technical refinement of (7).

Assumption 3 A. is imposed for simplicity of formulation. However, since g is determined by ψ, a known density, and by f, the density of interest, it is more natural to formulate the conditions in terms of f and ψ. If f is bounded, it is simple to see that g is also bounded since ψ is integrable. Further, suppose that the derivative of $f^{1/2}$ is bounded. Since under sufficient regularity,

$$g'(x) = \int_{\mathbb{R}} f'(x-y)\psi(y)\, dy,$$

and from the Schwarz inequality,

$$g^{1/2}(x) \geq \int_{\mathbb{R}} f^{1/2}(x-y)\psi(y)\, dy,$$

it follows that $g'/g^{1/2}$ is also bounded. Finally, there are several conditions on f and ψ that guarantee that g is bounded away from 0 on $[0,1]$. If ψ is bounded away from 0 on some interval $[-\epsilon,\epsilon]$ for $\epsilon > 0$, this will follow if f is bounded away from 0 on $[0,1]$. In this case, which includes e.g. the case of Laplace deconvolution, one can simply replace g by f in Assumption 3 A. However, if ψ has support on $(0,\infty)$, like e.g. the exponential distribution, one requires that f is bounded away from 0 on some interval $[-\epsilon,1]$.

Finally, Assumption 3 B. is a smoothness assumption on f. It implies in particular that f is p-times continuously differentiable, and it is satisfied if f is more than p-times differentiable.

Example 1 *(continued).* If Φ_ψ is the reciprocal of a polynomial as in Example 1, it is easy to check that (13) holds, and therefore Assumption 2 A. is satisfied. Furthermore we evidently have $C_\epsilon = 1/a_\beta$. Since $\Phi_\psi(-t) = \overline{\Phi_\psi(t)}$, it is clear that if β is even, $a_\beta \in \mathbb{R}$, while if β is odd, $a_\beta \in i\mathbb{R}$. Therefore the leading terms in (14) and (15) cancel out, and only terms of order $O(h)$ remain. Thus Assumption 2 B. holds true as well.

4 Bootstrap confidence bands

In this section we construct bootstrap confidence bands for the deconvolution problem. We consider the simple nonparametric bootstrap, where we resample n-times from the observations $X_1, \ldots, X_n$, and in this way obtain an i.i.d. sample $X_1^*, \ldots, X_n^*$ with distribution G_n, the empirical distribution function of the $X_1, \ldots, X_n$.

In a direct density estimation problem this method was studied by Hall (1993) via the Edgeworth expansion. Hall (1993) concluded that bootstrap confidence bands perform much better than asymptotic bands. Here we give a simple argument to show that the nonparametric bootstrap also works to construct confidence bands in the deconvolution problem. We denote by E^* and P^* conditional expectation and conditional probability, given $X_1, \ldots, X_n$. The bootstrap estimator of $f^{(j)}$ is given by

$$\hat{f}_n^{*(j)}(x) = \frac{1}{nh^{j+1}} \sum_{k=1}^{n} K_n^{(j)}\Big(\frac{x - X_k^*}{h}\Big). \tag{16}$$

Since the empirical characteristic function $\hat{\Phi}_n^*$ of $X_1^*, \ldots, X_n^*$, satisfies $E^*\hat{\Phi}_n^*(t) = \hat{\Phi}_n(t)$, using formula (3) we see that

$$E^* \hat{f}_n^{(j)*}(x) = \hat{f}_n^{(j)}(x).$$

The bootstrap approximation of $Y_n^{(j)}(t)$ is given by

$$Y_n^{*(j)}(t) = \frac{n^{1/2} h^{\beta+j+1/2}}{\tilde{g}_n(t)^{1/2}} \big(\hat{f}_n^{*(j)}(t) - \hat{f}_n^{(j)}(t)\big).$$

In order to construct bootstrap confidence bands, we simulate the distribution $\|Y_n^{*(j)}\|_{[0,1]}$. Let $q_{1-\alpha}^*$ denote its $1-\alpha$ quantile. In this way we obtain the bootstrap level-α confidence

band,

$$\hat{f}_n(t) - \frac{q^*_{1-\alpha}\tilde{g}_n^{1/2}(t)}{\sqrt{n}h^{\beta+j+1/2}} \le f(t) \le \hat{f}_n(t) + \frac{q^*_{1-\alpha}\tilde{g}_n^{1/2}(t)}{\sqrt{n}h^{\beta+j+1/2}}, \quad t \in [0,1]. \tag{17}$$

The following theorem shows that this method works asymptotically. Let d denote the Prohorov metric on the space of probability measures on $\mathbb{R}$. The precise definition is not relevant for us, we mainly need that it is a metric for weak convergence of probability measures on $\mathbb{R}$. Further, if U and V are real-valued random variables and $a \in \mathbb{R}$, $d\big(\mathcal{L}(U), \mathcal{L}(V)\big) = d\big(\mathcal{L}(U-a), \mathcal{L}(V-a)\big)$. Here $\mathcal{L}$ denotes the law (distribution) of a random variable.

Theorem 3. *Suppose that* $\|\tilde{g}_n - g\|_{[0,1]} = o\big((\log(1/h))^{-1}\big)$ *a.s.. Under Assumptions 1 – 3 and the bandwidth assumptions of Corollary 2, setting* $c_n = \big(2\log(1/h)\big)^{1/2}$ *we have that*

$$d\Big(\mathcal{L}(c_n\|Y_n^j\|_{[0,1]}), \mathcal{L}^*(c_n\|Y_n^{*(j)}\|_{[0,1]})\Big) \to 0 \quad a.s., \tag{18}$$

where $\mathcal{L}^*$ *is the conditional law given* $(X_i)_{i\ge1}$*. Therefore,*

$$P^*\Big(c_n\big(\|Y_n^{*(j)}\|_{[0,1]}/C_{K,1}^{1/2} - d_n\big) < x\Big) \to \exp(-2\exp(-x)) \quad a.s.,$$

where the constants d_n *and* $C_{K,1}$ *are as in Theorem 1.*

The almost sure uniform convergence $\|\tilde{g}_n - g\|_{[0,1]} = o\big((\log(1/h))^{-1}\big)$ of $\tilde{g}_n$ to g on the compact interval $[0,1]$ is e.g. satisfied for kernel density estimators with bandwidth h if the kernel is of bounded variation and has compact support, if f and hence also g is differentiable, c.f. Stute (1982). Note that this condition in particular implies the convergence in probability (11). Further observe that without the bandwidth assumption of Corollary 2, Theorem 3 yields the consistency of the bootstrap for constructing confidence bands for the smoothed version $E\hat{f}_n^{(j)}$ of $f^{(j)}$.

Remark 2. Neumann and Pohlzehl (1998) construct bootstrap confidence bands in a regression context by using the so-called wild bootstrap. They use rather sophisticated arguments to construct a version of the bootstrap process which approximates the original process directly, not only in distribution as in (18). However, in our context the approximation in distribution appears to be sufficient and in addition is much simpler, since we can apply the theory of empirical processes.

Further, Neumann and Pohlzehl (1998) do not require a limit theorem like Theorem 1 for their process (and in fact are uncertain about the limiting distribution of their statistics). Instead, they only derive a lower bound on the probability that the supremum statistic of the bootstrap process falls into small intervals. In addition with their bootstrap approximation result, this is enough to obtain valid bootstrap confidence bands. Such a technique might be useful when studying less regular deconvolution problems, which satisfy (6) but not necessarily (7).

5 Simulation results

5.1 Simulation framework

In this section we investigate the methods discussed above in a simulation study. To this end we generate data from the model (1). We consider two different densities for Z. The first one is the normal density

$$f_1(x) = \phi_{0.5,0.09},$$

where $\phi_{a,b}$ denotes the density of a normal random variable with mean a and variance b, and is therefore an infinitely smooth density. The second density represents the case of finite smoothness, and is given by

$$f_2(x; k, \theta) = x^{k-1} \frac{\exp(-x/\theta)}{\theta^k \Gamma(k)}, \text{ for } x > 0,$$

i.e. a Gamma distribution, where we fix the shape parameter to $k = 6$ and the scale parameter to $\theta = 0.08$. We assume the error terms ϵ_i to be independent from the Z_i and have Laplace density $\psi(x; \lambda) = (\lambda/2) \exp(-\lambda|x|)$, $x \in \mathbb{R}$, where $\lambda = 3$ for the simulations involving the underlying density f_1, and $\lambda = 10$ for those which involve f_2. This amounts to "signal-to-noise" ratios $\sqrt{\mathrm{Var}(f_1)/\mathrm{Var}(\psi)} = 0.6$ and $\sqrt{\mathrm{Var}(f_2)/\mathrm{Var}(\psi)} = 1.4$, respectively. Moreover, we will perform some simulations with underlying density f_2 and Gaussian noise of variance $\sigma^2 = 0.02$, i.e. with the same variance as a Laplace density with parameter $\lambda = 10$.

Our procedure to determine confidence bands for the density f of Z is as follows. First, we estimate f by the estimator (3), based on the compactly supported flattop kernel with Fourier

Figure 1

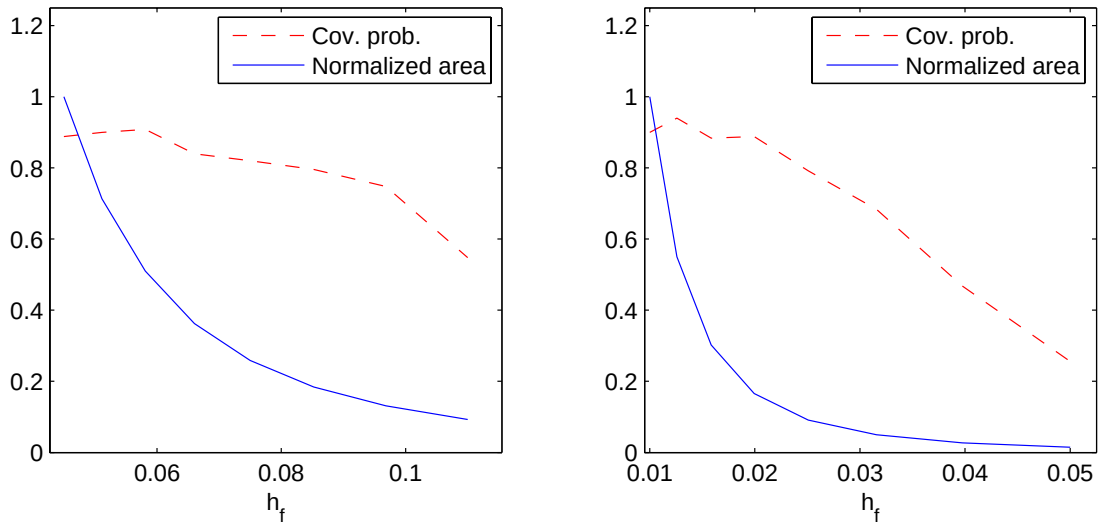

Figure 1: Coverage probability and area of asymptotic confidence bands for estimation of the densities f_1 (left panel) and f_2 (right panel) from 500 observations. The area of the confidence bands has been determined by numerical integration and is normalized in this plot to a maximum value 1 in both panels. The nominal coverage probability of the confidence bands is 90%.

Figure 2

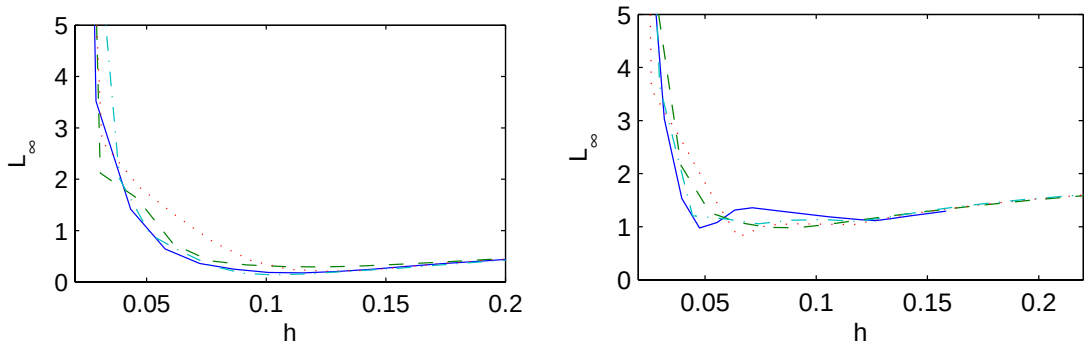

Figure 2: L_∞-distance between the true densities f_1 (left) and f_2 (right) and estimates $\hat{f}_{n,h}$ in dependence of the bandwidth. The samples in the simulations consist of 500 observations contaminated by Laplace distributed noise with $\lambda = 3$ for simulations involving f_1 and $\lambda = 10$ for those involving f_2, respectively. The four different lines correspond to four random datasets each for estimation of f_1 respectively f_2.

transform

$$\Phi_k(t) = \begin{cases} 1 & \text{if } |t| \leq c \\ \exp \frac{-b \exp(-b/(|t|-c)2)}{(|t|-1)2} & \text{if } c < |t| < 1, \\ 0 & \text{if } |t| \geq 1 \end{cases}$$

with $b = 1$ and $c = 0.05$ (Politis and Romano, 1999) and bandwidth h_f. In our simulations (not displayed) it turned out that this kernel results in better, more stable, numerical performance of $\hat{f}_n^{(j)}$ than the sinc-kernel with Fourier transform $\Phi(t) = 1_{[-1,1]}(t)$. This is mainly due to the fact, that $\Phi_k(t)$ is both infinitely differentiable and flat near the origin. Then we construct a confidence band in the interval $[0, 1]$ according to the methods discussed in Theorem 1 for the asymptotic confidence bands and along the lines described in Section 4 (cf. (17)) for the bootstrap confidence bands, respectively. Here we estimate the density g of X with a kernel density estimator with Gaussian kernel ϕ_{0,h_g^2} of bandwidth h_g.

5.2 Bandwidth selection

We will now discuss selection of the bandwidths h_f and h_g. The more easy part is to choose the bandwidth h_g for "direct" density estimation of g due to the additional smoothness imposed

Figure 3

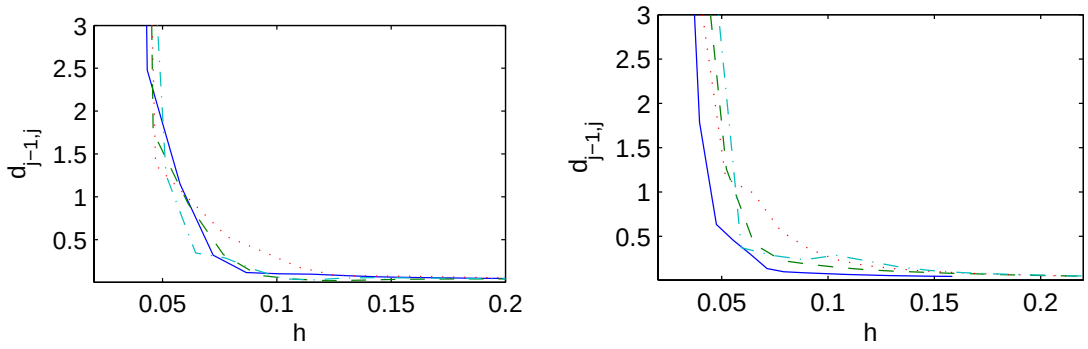

Figure 3: L_∞-distance $d_{j,j+1}^{(\infty)}$ between estimates of f_1 (left) and f_2 (right), respectively, in dependence of the bandwidths h_j, using the same line style for the same random datasets (and hence estimates) as in Fig. 2.

Table 2

Nominal cov. (%)	Method	Cov. prob.	Area	Cov. prob.	Area
		Gaussian distribution		Gamma distribution	
80	Bootstr.	81.6	0.95	90.8	0.60
90	Bootstr.	88.0	1.10	97.6	0.68
95	Bootstr.	90.8	1.23	98.8	0.74

Table 2: Coverage probabilities and confidence band areas for estimating the Gaussian density f_1 and the density of a Gamma distribution f_2 from 500 observations, contaminated by Laplace distributed noise with variance $2/9$ for simulations involving f_1 and variance 0.02 for those involving f_2. The bandwidth h for the estimator $\hat{f}_{n,h}$ was selected in a fully data-driven way. See text for details.

Table 3

Nominal cov. (%)	Method	Cov. prob.	Area	Cov. prob.	Area
		Gamma dist, $J = 50$		Gamma dist, Lap. $*$ Lap.-noise	
80	Bootstr.	86.0	0.37	90.4	0.65
90	Bootstr.	90.0	0.42	94.8	0.75
95	Bootstr.	93.6	0.46	97.2	0.84

Table 3: Coverage probabilities and confidence band areas for estimating the density of a Gamma distribution f_2 from 500 observations, contaminated by Laplace noise and with $J = 50$ (left) and by noise distributed according to a Laplace distribution with variance 0.02 convolved with itself (right). See text for details.

on g by the convolution with ψ. We found in our simulations that the coverage probabilities and confidence band areas are quite insensitive to the value of h_g and any of the standard methods will be sufficient, e.g. cross validation or, where it is reasonable, a normal reference estimator, as used in our simulations. However, h_f has to be selected more carefully. This is illustrated in Fig. 1, which shows the coverage probabilities and area for 90% nominal coverage confidence bands as a function of the bandwidth h_f used in the estimator (3). In this simulation the true density is f_1 and f_2, respectively, and in each case we performed 250 simulations with a sample size of 500, respectively. We conclude from Fig. 1 that for large bandwidth the confidence bands are narrow, but the true coverage probability is far below its nominal level. This is due to the (uncorrected) bias, which increases with the bandwidth. For small bandwidths, we undersmooth the estimator, and the true coverage probability approaches its nominal value. Hence, as a major conclusion, we see from this that undersmoothing is advisable in order to maintain the coverage probability of the confidence bands. However, at the same time this leads to an increase in the area of the confidence bands. Hence, choosing the bandwidth h_f implies a trade-off between meeting the nominal coverage probabilities and reducing the area of the confidence bands.

In general, in a practical application we would not necessarily recommend selecting the bandwidth such that it is *undersmoothing*, but certainly such that it is *not oversmoothing*. This is

somewhat in contrast to estimating the function f itself, where slight oversmoothing typically still recovers the main structural features of f, such as peaks, valleys and modes. In fact, several approaches to data-driven bandwidth selection for deconvolution estimators have been investigated in the past, mainly with the purpose to minimize the mean integrated square error as optimality criterion. These include cross-validation procedures (Stefanski & Carroll, 1990; Hesse, 1999), bootstrap methods (Delaigle & Gibjels, 2001), and methods based on plug-in and to some extend SEQ (solve-the-equation) ideas (Delaigle & Gijbels, 2004). The bandwidth selected by these methods are in general oversmoothing and in some cases show a large scatter (cf. Delaigle & Gijbels, 2004, Figs. 2, 4, 6, and 7). Therefore it is not straightforward to devise a general rule, how the bandwidth from one of the previously mentioned methods could be adapted to our requirements directly.

An L_∞-based bandwidth selector. Since the uniform confidence bands considered here are based on an analysis of the L_∞-distance between estimates $\hat{f}_{n,h}$ and f, we therefore suggest in the following a bandwidth selection procedure which aims at selecting the bandwidth h_{opt} which minimizes the L_∞-distance between $\hat{f}_n$ and the underlying density f as surrogate optimality criterion.

In the following we pursue an approach based on estimation of the bandwidth h^{opt}, which minimizes the L_∞-distance $d^{(\infty)}(f, \hat{f}_{n,h})$ between the underlying density f and the estimator $\hat{f}_{n,h}$ for bandwidth h. Let us mention that confidence bands based on h_{opt} would by Theorems 1 and 3 only be valid for $E\hat{f}_n^{(j)}$, since the bias remains uncorrected. However, our simulation study shows that in finite samples, the subsequent bandwidth selection strategy yields satisfactory results for $f^{(j)}$ itself in finite samples.

Fig. 2 shows for four different samples the L_∞-distance between estimates $\hat{f}_{n,h}$ and the underlying densities f_1 and f_2, respectively, in dependence of h. For oversmoothing bandwidths, i.e. $h > h^{\mathrm{opt}}$ the estimators change only moderatly with increasing bandwidth, to the effect, that the extrema of the estimated density increasingly get smoothed out. In consequence, the L_∞-distance changes only slowly. The situation is different for undersmoothing bandwidth $h < h^{\mathrm{opt}}$. With decreasing bandwidth, those frequencies in the spectral domain where the empirical characteristic function is dominated by noise in the data contribute more and more

to the estimator, producing increasingly strong artificial oscillations. This results in a sudden steep increase of $d^{(\infty)}(f, \hat{f}_{n,h})$ for decreasing bandwidth $h < h^{\text{opt}}$. In consequence, $h \approx h^{\text{opt}}$ (or slightly smaller) is a straightforward choice for the bandwidth for generating the confidence bands. This is, however, not accessible directly because the criterion $d^{(\infty)}(f, \hat{f}_{n,h})$ to be evaluated depends on f itself. In the following this minimum will therefore be estimated. We have argued in the previous paragraph that $\hat{f}_{n,h}$ changes in a different way with bandwidth h for $h < h^{\text{opt}}$ and $h > h^{\text{opt}}$. In Fig. 3 we show for the same simulations as in Fig. 2 the L_∞-distance

$$d^{(\infty)}_{j-1,j} := d^{\infty}\left(\hat{f}_{n,h_j}, \hat{f}_{n,h_{j-1}}\right).$$

between estimators f_{n,h_j} for two adjacent bandwidths of a sequence of bandwidths $h_j = h_0 \cdot (j/20)$, $j = 1, \ldots, 20$, where h_0 is some oversmoothing pilot bandwidth. A comparison of Figs. 2 and 3 indicates that the bandwidth, where $d^{(\infty)}_{j-1,j}$ changes its slope suddenly, is a good indicator of the bandwidth with minimizes the L_∞-distance $d^{(\infty)}(f, f_{n,h})$.
We therefore suggest the following two-step algorithm for a data-driven bandwidth selection.

Initialization step: Estimate a pilot bandwidth h_0 which is oversmoothing. If the density to be estimated is expected to be unimodal, a normal reference bandwidth estimator (cf. Delaigle & Gijbels, 2004) is sufficient, however in general a more sophisticated selection algorithm such as the plug-in estimator of Delaigle & Gijbels (2004) is to be preferred. To make sure that the pilot bandwidth is not undersmoothing it could be increased by a factor $\gamma > 1$, say $\gamma = 1.5$. For a guidance on how to select γ, compare, e.g. Figs. 2,4,6 and 7 in Delaigle & Gijbels (2004).

Selection step: Compute $\hat{f}_{n,h_j}$ for a grid of J values $h_j = j \cdot (h_0/J)$, $j = 1, \ldots, J$, where $J \approx 20$ is found to be sufficient in our case. Choose the largest bandwidth h^* such that the "change between models for adjacent bandwidth" $d^{(\infty)}_{j-1,j}$ is more than τ times ($\tau > 1$) larger than $d^{(\infty)}_{J-1,J}$ in the case of the pilot bandwidth.

There are several parameters involved in this bandwidth selection algorithm, and we now comment on the sensitivity of the resulting bandwidth when these are varied. Simulations with a number of different values for γ showed that its precise choice does not impact the

results for the bandwidth significantly. Most importantly here is to guarantee an initial bandwidth h_0 which is oversmoothing. Fine-tuning of the threshold parameter τ allows to adapt the method to different possible purposes of the estimator $\hat{f}_{n,h}$ - decreasing τ yields larger bandwidths and in consequence smoother estimates of f. In our simulations $\tau \approx 2$ turned out to be a good choice for the confidence bands, *independent of sample size and "true" underlying density* f. This is illustrated by Table 2, which gives the results of a small simulation study of the bootstrap confidence bands consisting of 250 simulations with 200 bootstrap samples each. The confidence bands perform remarkably well for the estimates based on our bandwidth selection algorithm.

Additional simulations where performed to test for a dependence of the estimated bandwidth on the bin number J of the bandwidth selection algorithm, and on the noise density (in particular the resulting degree of ill-posedness of the deconvolution problem). Table 3 reports the results of these simulations. In the first part, we have modified the number J considered for the selection of the bandwidth, which yields a slight improvement in the performance of the confidence bands, particularly w.r.t. the mean area they cover. This is due to the fact that now more different possible choices for the bandwidth h_j are available to select from, and the critical bandwidth h^* can be determined more precisely. In consequence, it happens less frequently in the simulations that a significantly too small bandwidth is chosen, which would produce strong oscillations in the estimator $\hat{f}_{n,h}$, and eventually large confidence bands. As second test we replaced the Laplace distribution of the noise terms by a Laplace convolved with itself in order to increase the ill-posedness of the problem by changing β from 2 to 4. Again, the results in Table 3 show that our bandwidth choice is robust against this change in the degree of ill-posedness.

The suggested bandwidth selection procedure requires the computation of estimates $\hat{f}_{n,h}$ for a grid of different bandwidths h_j. In addition to the fully automatic bandwidth selection as described above, it is recommendable to further examine these estimates visually, if possible. In some preliminary simulations it turned out that the performance of the confidence bands can indeed be improved further by visual examination of the estimates $\hat{f}_{n,h_j}$ over a range of bandwidth $h_k, \ldots, h_l$, $1 \leq k \leq l \leq J$, and selection of the smallest amongst these bandwidths

Table 4

Nominal cov. (%)	n	Method	Cov. prob.	Area
80	100	Asympt.	56.4	0.71
		Bootstr.	74.1	0.80
	500	Asympt.	58.4	0.41
		Bootstr.	77.2	0.50
90	100	Asympt.	68.8	0.84
		Bootstr.	88.8	0.95
	500	Asympt.	70.8	0.49
		Bootstr.	87.2	0.59
95	100	Asympt.	80.4	0.96
		Bootstr.	93.3	1.07
	500	Asympt.	80.8	0.55
		Bootstr.	94.0	0.66

Table 4: Coverage probabilities and confidence band areas for estimation of the Gaussian f_1 from 100 and 500 observations, respectively, contaminated by Laplace distributed noise with variance 2/9.

for which the estimate does not show significant oscillations to negative values. We remark that this approach is closely related to the scale space view of curve estimation (e.g. Chaudhuri and Marron, 2000), where one examines the empirical scale space surface generated by $\hat{f}_h(x)$ as function of h, x for significance of present features such as peaks or valleys. A further investigation of the theoretical properties of data-driven bandwidth estimation from a scale space point of view for confidence bands in deconvolution density estimation would clearly be of some interest, but is beyond the scope of this paper.

5.3 Comparison of asymptotic and bootstrap confidence bands

In order to investigate the performance of the bootstrap confidence bands in comparison to the asymptotic case (cf. Corollary 2) in the following simulations we estimated in a first

Table 5

Nominal cov. (%)	n	Method	Cov. prob.	Area
80	100	Asympt.	33.6	0.97
		Bootstr.	80.0	1.06
	500	Asympt.	48.8	0.94
		Bootstr.	84.0	0.49
90	100	Asympt.	45.2	1.08
		Bootstr.	86.0	1.21
	500	Asympt.	58.8	1.04
		Bootstr.	94.0	0.55
95	100	Asympt.	53.6	1.20
		Bootstr.	92.0	1.30
	500	Asympt.	70.4	1.14
		Bootstr.	98.0	0.59

Table 5: Coverage probabilities and confidence band areas for estimation of the density of the Gamma distribution f_2 from 100 and 500 observations, respectively, contaminated by Laplace distributed noise with variance 0.02.

Table 6

Nominal cov. (%)	n	Method	Cov. prob.	Area	Cov. prob.	Area
			correct specification		misspecification	
80	100	Bootstr.	88.0	0.62	78.8	0.50
	500	Bootstr.	87.2	0.37	82.4	0.36
90	100	Bootstr.	94.4	0.72	86.4	0.58
	500	Bootstr.	93.2	0.43	90.8	0.41
95	100	Bootstr.	96.8	0.80	92.8	0.65
	500	Bootstr.	96.0	0.48	94.8	0.46

Table 6: Coverage probabilities and confidence band areas for estimating the density of the Gamma distribution f_2 from 100 and 500 observations, respectively, contaminated by Gaussian noise with variance 0.02. The first two columns show the results for correctly specified noise, while the last two columns show those for a misspecified noise model, where the error is assumed to be Laplace distributed with variance 0.02.

step h^{opt} by minimizing $d^{\infty}(f_i, \hat{f}_{n,h})$, $i = 1, 2$, for several samples and subsequent averaging. This procedure was performed for every combination of sample size 100, 500 with underlying density f_1 and f_2, respectively. The bandwidth was then kept fixed in the subsequent simulations. In the first part of our simulation study, we estimate the density f_1 from samples of sizes 100 and 500 generated according to model 1. We perform 250 simulation runs, and for each simulation of bootstrap confidence bands, 200 bootstrap samples are generated. Table 4 presents the results for nominal coverage probabilities 80%, 90% and 95%. We simulated coverage probabilites together with the area of the confidence bands for the estimator $\hat{f}_{1,n}$ of f_1. From the table we observe that both for the asymptotic and the bootstrap confidence bands the results improve with increasing sample size. In a second Monte Carlo study we have used the Gamma density f_2. Here, we focus on the confidence intervals for $0.25 < x < 0.7$, where the density departs significantly from 0. Apart from this we use the same simulation setup as described above for the case of f_1. Table 5 summarizes the results for asymptotic and bootstrap confidence bands. Both methods produce reasonable results for the coverage

probabilities with a clear advantage for the bootstrap bands. However, in comparison to the simulations based on f_1, the nominal coverage probabilities of the asymptotic confidence bands are recovered less precisely for f_2. This difference in the quality of approximation is somewhat reflected by the difference in the rates of convergence of the function estimates $\hat{f}_1$ and $\hat{f}_2$, which are $O\left(\log(n)^{5/4}/\sqrt{n}\right)$ (Butucea, 2004) and $O(n^{-1/3})$ (e.g. Fan, 1991a), respectively. However, Fig. 4, which presents some typical examples for estimated asymptotic confidence bands from 500 and 5000 simulations, shows that these still produce reasonable results for moderate sample sizes such as $n = 500$. Finally, when comparing these results based on the optimal bandwidths h^{opt} to those obtained for our data–driven selection procedure, one observes that the bootstrap confidence bands perform slightly better with $h = h^{\text{opt}}$ than if the bandwidth is chosen in a data–driven way. In particular, the nominal coverage probabilities are met to comparable precision for both bandwidth choices, but the areas of the bands are smaller for $h = h^{\text{opt}}$.

5.4 Robustness and misspecification of ψ

In many practical situations the true shape of the density ψ of the noise is known only approximately and a polynomial decay of Φ_ψ as in assumption (6) may not be satisfied. In particular, Gaussian deconvolution is not covered by our theory and it might be of interest how the bootstrap confidence bands in (17) perform in this case. An example from astrophysics will be discussed below in Section 6. The following simulation study indicates that the proposed bootstrap estimator of the confidence bands is robust against this kind of noise model misspecification and still performs well (cf. Hesse, 1999; Meister, 2004). To this end we have performed two additional sets of simulations with f_2 as underlying density and the same simulation setup as used so far. Now the noise term ε follows a normal density with variance $\sigma^2 = 0.02$, i.e. with the same variance as in our previous simulations. Note, that it is not possible to use asymptotic confidence bands as in (12) anymore, however, we still can implement their bootstrap version.

In the first part of the simulations we assume that the true noise density is known to be $\phi_{0,0.02}$. The results are summarized in the first two columns of Table 6. Finally, in the second

part, we assume in the definition of the deconvolution estimator $\hat{f}_n$ that the noise density is Laplace distributed with variance 0.02, i.e. we misspecify the noise model where we assume the same variance in order to keep the results comparable. Results for this case are given in the two rightmost columns of Table 6. From a comparison of the results in Table 5 with those in Table 6 we conclude that the bootstrap estimator for the confidence bands is reasonably robust against misspecification of the noise model and also performs well for the exponentially ill-posed case of Gaussian noise.

The robustness indicated in our simulations fits well with theoretical results of Meister (2004). He analyzed the effect of misspecification of the error density in density deconvolution on the asymptotic behaviour of the mean integrated squared error [MISE] and found that if the true error density is normal, but misspecified as Laplace in the density estimator $\hat{f}_n$, the MISE does still possess an upper bound. On the other hand, Meister (2004) shows that if the true error density is Laplace, but misspecified as normal in the setup of $\hat{f}_n$, then the MISE diverges to infinity. He therefore recommends to select the Laplace density to set up $\hat{f}_n$, if a choice between a normal and a Laplace distributed error density has to be made.

5.5 Summary

From the simulation study we have found that both the asymptotic and the bootstrap confidence band yield reasonable results if interactively post-processed bandwidths are used, with some advantage for the bootstrap method, particularly for underlying densities of finite smoothness. This is in accordance with Hall's (1993) findings for the direct case. Moreover, the bootstrap confidence bands seem to be reasonably robust against a misspecified noise model, notably, if the deconvolution density leads to a severly ill-posed problem. In practical applications, if visual selection of the bandwidth is not possible or desirable to avoid any subjectiveness of the estimate, the data-driven bandwidth selectior with bootstrap confidence bands provides a sound method, which is rather insensitive to the precise choice of the parameters γ, τ, N of the algorithm, and to the underlying densities f and ψ.

In conclusion, as it is well-known from indirect density estimation and regression, bandwidth selection has to be done very carefully, and appears to be in general more difficult than in

Figure 4

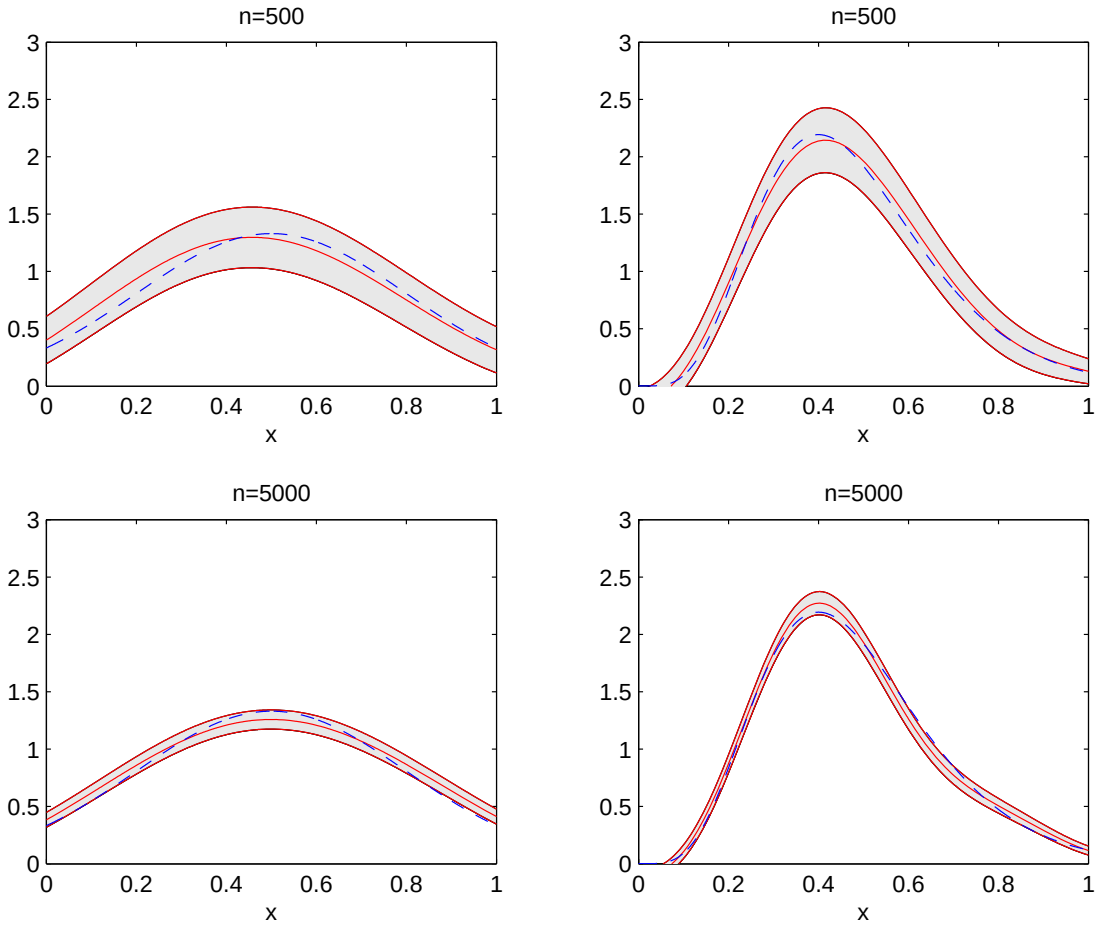

Figure 4: Estimate $\hat{f}_n$ and associated 90% nominal coverage probability asymptotic confidence bands (solid lines) of the true densities f_1 (l.h.s., dashed lines) and f_2 (r.h.s., dashed lines) from 500 and 5000 observations, respectively.

the direct case.

6 Metallicity distribution of F and G dwarfs in the Solar neighbourhood

The apparent lack of metal-poor G dwarfs relative to predictions from "closed-box models" of stellar formation (the so-called "G dwarf problem") is an interesting observation that has to be explained by stellar formation theory. In this section we apply our method to a sample of approximately 1900 stellar metallicities from the Geneva-Copenhagen survey of kinematic properties, ages and metallicities of F and G dwarfs in the Solar neighbourhood (Nordström *et al.*, 2004).

Metallicities such as $Z := [Fe/H]$, i.e. the log-based relative fraction of iron compared with hydrogen, are a measure of the fraction of heavy elements in the star. Since the formation of the universe stars have continuously enriched the interstellar gas, out of which stars form, with metals. Stars of different age thus contain different amounts of metals, which makes the latter quantity of significant interest to understand the formation history of the Milky Way and to test models of stellar formation (Pagel, 1997).

It is not possible to measure stellar metallicities in situ. Instead, they can be derived from observations of the brightness of a star in certain spectral bands using calibrations, which are rather difficult to determine (e.g. Schuster and Nissen, 1989; Edvardson *et al.*, 1993; Chen *et al.*, 2000). We assume that the error made in this conversion is to a reasonable extend of stochastic nature, as indicated e.g. by Fig. 8 in Nordström *et al.* (2004), and can be estimated from the dispersions σ between the metallicities derived for the same star from different calibrations. For our data typical values are $\sigma \approx 0.1$ (Nordström *et al.*, 2004), which is the value we use below. We assume further that the observed metallicities $X_1, \ldots, X_n$ may be modelled as

$$X = [Fe/H] + W,$$

where the noise W is distributed according to a centered Laplace distribution with variance σ^2, and $[Fe/H]$ is stochastically independent of W. A standard assumption in astronomy

Figure 5

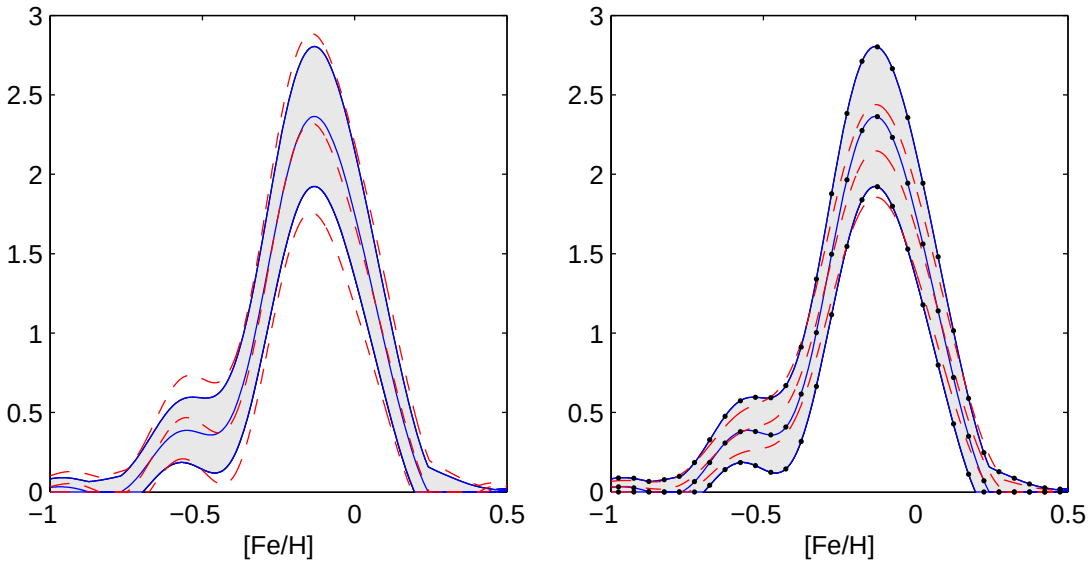

Figure 5: Estimates and associated 90% nominal coverage probability bootstrap confidence bands of the metallicity density of F and G dwarfs in the solar neighbourhood. Left figure: estimate and confidence bands for $\hat{f}_n$ based on Laplace distributed noise with variance 0.1^2 (solid lines and shaded area), and estimate and confidence bands for $\hat{f}_n$ based on normal noise with variance 0.1^2 (dashed lines). Right figure: estimates and associated confidence bands for $\hat{f}_n$ based on Laplace distributed noise with variance 0.08^2 (dashed lines), 0.1^2 (solid lines) and 0.12^2 (lines with dots), respectively. Note that the confidence bands in the case of variance 0.12^2 nearly coincide with those for variance 0.1^2.

would also be that the noise follows a Gaussian law. However, in general, the distribution of the noise is not known very well, and as we have seen in section 5 both a Laplace and a Gaussian density with a certain variance lead to similar results. In accordance with these results, and with the recommendations of Meister (2004) we use a Laplace density to set up the density estimator $\hat{f}_n$. However, to facilitate a further comparison of this approach, the figure also shows confidence bands for the estimator $\hat{f}_n$ based on a normal noise assumption with the same variance $\sigma^2 = 0.1^2$ as used in the case of Laplace distributed noise. Moreover, since the variance of the noise is known only approximately, we have also determined confidence bands for $\hat{f}_n$ for a Laplace distributed noise with variance 0.08^2 and 0.12^2, respectively. We note that in all these computations we followed the procedure described in section 5 to select the bandwidth.

Fig. 5 presents the estimators $\hat{f}_n$ of the density of $[Fe/H]$ and the associated confidence bands with nominal coverage probility 90% from 500 bootstrap replications for the different setups of the density estimator $\hat{f}_n$. From the figure we conclude that neither the misspecification considered for the error density in the setup of $\hat{f}_n$, nor a reasonable amount of misspecification of the noise variance do change the result strongly.

A comparison of our results with Fig. 7 in Sommer-Larsen (1991) confirms the "G dwarf problem" in "closed-box models" models of stellar formation, i.e. they predict significantly too many F- and G-dwarfs of low metallicity ($[Fe/H] \leq -0.6$). However, a comparison with Figs. 16 and 17 in Jørgensen (2000) shows that the theoretical infall models of Lynden-Bell (1975) and Pagel and Tautvaisiene (1995) reproduce the F and G dwarf metallicity distribution approximately within the confidence bands, which gives further evidence to these models.

Acknowledgements

We would like to thank Jon Wellner for helpful discussions, and the associate editor and two anonymous referees for helpful comments. Financial support of the Deutsche Forschungsgemeinschaft, Grant MU 1230/8-1, and of SFB 475 is gratefully acknowledged.

Appendix

Proof of Theorem 1. Following Bickel and Rosenblatt (1973), we will prove Theorem 1 by approximating $Y_n^{(j)}$ by a Gaussian process which does not depend on the density f, and then apply a result for the distribution of the supremum of a Gaussian process. Some steps of the proof are similar to those in Piterbarg and Penskaya (1993), who use the method of Bickel and Rosenblatt to determine the asymptotic distribution of the integrated squared error. However, in order to obtain the asymptotic distribution of supremum-type statistics, finer approximations are required (cf. Lemmas 4 and 5).

To keep the proof more transparent we have split it into several lemmas. Let G denote the distribution function of the X_i. W.l.o.g. we can assume that $X_i = G^{-1}U_i$, where the U_i are i.i.d. uniform on $[0,1]$, and $G^{-1}(q) = \inf\{t \in \mathbb{R} : \ G(t) \geq q\}$. Let G_n be the empirical distribution function of $X_1, \ldots, X_n$, and let $\alpha_n^G(t) = \sqrt{n}\big(G_n(t) - G(t)\big)$ be the empirical process. From Komlós *et al.* (1975) we can assume that there exist Brownian bridges B_n such that

$$\sup_{t\in\mathbb{R}} |\alpha_n^G(t) - B_n(G(t))| = O_P\Big(\log(n)/\sqrt{n}\Big).$$

Further $B_n(t) = W_n(t) - tW_n(1)$, where W_n are Wiener processes. Partial integration and the fact that $K_n^{(j)}(x) \to 0, |x| \to \infty$ gives

$$\begin{aligned} Y_n^{(j)}(t) &= \frac{h^{\beta-1/2}}{g(t)^{1/2}} \int_{\mathbb{R}} K_n^{(j)}\Big(\frac{t-x}{h}\Big)\, d\alpha_n^G(x) \\ &= \frac{h^{\beta-3/2}}{g(t)^{1/2}} \int_{\mathbb{R}} K_n^{(j+1)}\Big(\frac{t-x}{h}\Big)\, \alpha_n^G(x)\, dx \quad t \in [0,1]. \end{aligned} \tag{19}$$

As approximations to the process $Y_n^{(j)}$ we consider the processes

$$Y_{n,0}^{(j)}(t) = \frac{h^{\beta-1/2}}{g(t)^{1/2}} \int_{\mathbb{R}} K_n^{(j)}\Big(\frac{t-x}{h}\Big)\, dB_n(G(x))$$

and

$$Y_{n,1}^{(j)}(t) = \frac{h^{\beta-1/2}}{g(t)^{1/2}} \int_{\mathbb{R}} K_n^{(j)}\Big(\frac{t-x}{h}\Big)\, dW_n(G(x)).$$

From the partial integration formula for stochastic integrals (c.f. Øksendal, 1998, p.46) one obtains expressions for $Y_{n,0}^{(j)}(t)$ and $Y_{n,1}^{(j)}(t)$ which are analogous to (19).

Lemma 4. *Under Assumptions 1, 2 A. and 3 A.,*

$$\|Y_n^{(j)} - Y_{n,0}^{(j)}\|_{[0,1]} = O_P\big(h^{-1/2}n^{-1/2}\log(n)\big).$$

Proof. Using (19) and its analogue for $Y_{n,0}^{(j)}$, for $t \in [0,1]$ we get

$$|Y_n^{(j)}(t) - Y_{n,0}^{(j)}(t)| \leq \frac{h^{\beta-3/2}}{g(t)^{1/2}} \|\alpha_n^G - B_n(G)\|_{\mathbb{R}} \int_{\mathbb{R}} \big|K_n^{(j+1)}\Big(\frac{t-x}{h}\Big)\big|\, dx$$

Substituting $u = (t-x)/h$ and using Assumptions 2 A. and 3 A., we obtain the result. □

Lemma 5. *Under Assumptions 1 and 3,*

$$\|Y_{n,0}^{(j)} - Y_{n,1}^{(j)}\|_{[0,1]} = O_P\big(h^{\beta+j+1/2}\big).$$

Proof. We have

$$|Y_{n,0}^{(j)}(t) - Y_{n,1}^{(j)}(t)| = \frac{h^{\beta-1/2}}{g(t)^{1/2}}\,|W_n(1)|\,\Big|\int_{\mathbb{R}} K_n^{(j)}\Big(\frac{t-x}{h}\Big)\,g(x)\,dx\Big|$$

From Parseval's formula, we get

$$\begin{aligned}\Big|\int_{\mathbb{R}} K_n^{(j)}\Big(\frac{t-x}{h}\Big)\,g(x)\,dx\Big| &= \Big|\frac{1}{2\pi}\int_{\mathbb{R}} h\exp(iut)(ihu)^j\,\frac{\phi_k(-hu)}{\Phi_\psi(-u)}\,\overline{\Phi_\psi(u)\Phi_f(u)}\,du\Big| \\ &\le \frac{1}{2\pi}h^{j+1}\int_{\mathbb{R}}\Big|\Phi_f(t)\Big|\,dt.\end{aligned}$$

□

Next let W be a two-sided Wiener process on $\mathbb{R}$, and consider the processes

$$Y_{n,2}^{(j)}(t) = \frac{h^{\beta-1/2}}{g(t)^{1/2}}\int_{\mathbb{R}} K_n^{(j)}\Big(\frac{t-x}{h}\Big)\,g(x)^{1/2}\,dW(x)$$

and

$$Y_{n,3}^{(j)}(t) = h^{-1/2}\int_{\mathbb{R}} K^{(j)}\Big(\frac{t-x}{h}\Big)\,dW(x),$$

where $K^{(j)}$ is given in (8). Notice that the process $Y_{n,2}^{(j)}$ has the same finite-dimensional distributions as $Y_{n,1}^{(j)}$. Moreover, from a partial integration,

$$Y_{n,2}^{(j)}(t) = \frac{h^{\beta-1/2}}{g(t)^{1/2}}\int_{\mathbb{R}}\Big(K_n^{(j+1)}(u)\,g(t-hu)^{1/2} + hK_n^{(j)}(u)\frac{g'(t-hu)}{g^{1/2}(t-hu)}\Big)W(t-hu)\,du. \tag{20}$$

Lemma 6. *Under Assumptions 1 - 3 with δ as in Assumption 2 B.,*

$$\|Y_{n,2}^{(j)} - Y_{n,3}^{(j)}\|_{[0,1]} = O_P\big(h^{\min(1/2,\delta)}\big).$$

Proof. From (20) and a corresponding formula for $Y_{n,3}$ we get

$$\begin{aligned}|Y_{n,2}^{(j)}(t) - Y_{n,3}^{(j)}(t)| \le\ & \frac{h^{1/2}}{g(t)^{1/2}}\Big|\int_{\mathbb{R}} h^\beta K_n^{(j)}(u)\frac{g'(t-hu)}{g^{1/2}(t-hu)}W(t-hu)\,du\Big| \\ &+\frac{h^{-1/2}}{g(t)^{1/2}}\Big|\int_{\mathbb{R}} g(t-hu)^{1/2}\big(h^\beta K_n^{(j+1)}(u) - K^{(j+1)}(u)\big)W(t-hu)\,du\Big| \\ &+h^{-1/2}\Big|\int_{\mathbb{R}}\big(\frac{g(t-hu)^{1/2}}{g(t)^{1/2}} - 1\big)\,K^{(j+1)}(u)W(t-hu)\,du\Big|\end{aligned}$$

The claim follows from Assumptions 2 - 3 and the law of the iterated logarithm for the Wiener process (c.f. Karatzas and Shreve, 1991, p. 112). □

The proof of Theorem 1 now follows from Lemmas 4 - 6 and an application of Theorem A1 in Bickel and Rosenblatt (1973) to the process $Y_{n,3}^{(j)}$. □

Proof of Corollary 2. Consider the processes

$$Y_{n,4}^{(j)}(t) = \frac{n^{1/2}h^{\beta+1/2}}{\tilde{g}_n(t)^{1/2}}\big(\hat{f}_n^{(j)}(t) - E\hat{f}_n^{(j)}(t)\big), \quad Y_{n,5}^{(j)}(t) = \frac{n^{1/2}h^{\beta+1/2}}{\tilde{g}_n(t)^{1/2}}\big(\hat{f}_n^{(j)}(t) - f^{(j)}(t)\big).$$

From Theorem 1, $\|Y_n^{(j)}\|_{[0,1]} = O_P\big(\log(1/h)\big)$. Since

$$Y_n^{(j)}(t) - Y_{n,4}^{(j)}(t) = \frac{\tilde{g}_n^{1/2}(t) - g^{1/2}(t)}{\tilde{g}_n^{1/2}(t)} Y_n^{(j)}(t),$$

we conclude that $\|Y_n^{(j)} - Y_{n,4}^{(j)}\|_{[0,1]} = o_P\big((\log(1/h))^{-1/2}\big)$. Furthermore, from (10), uniformly for $t \in [0,1]$,

$$|Y_{n,4}^{(j)}(t) - Y_{n,5}^{(j)}(t)| = \frac{n^{1/2}h^{\beta+j+1/2}}{|\tilde{g}_n(t)^{1/2}|}\big|E\hat{f}_n^{(j)}(t) - f^{(j)}(t)\big| = O_P\big(n^{1/2}h^{\beta+r-1/2}\big).$$

Therefore the conclusion of Theorem 1 also remains valid for the process $Y_{n,5}^{(j)}$, and (12) follows from rearranging the terms. □

Proof of Theorem 3. We start by recalling an approximation to the bootstrap empirical process due to Shorack (1982). Let $U_1^*, U_2^*, \ldots$ be i.i.d. uniform on $[0,1]$, and let B_n^*, $n \geq 1$, be Brownian bridges, where both the U_i^* and the B_n^* are independent of the X_i, such that

$$\|\alpha_n^* - B_n^*\|_{[0,1]} = O_{P^*}\big(\log(n)/\sqrt{n}\big) \quad a.s., \tag{21}$$

where α_n^* is the empirical process of the U_i^*, and a.s. is with respect to the observations X_i. In the bootstrap procedure, w.l.o.g. we can assume that $X_i^* = G_n^{-1}U_i^*$, $i = 1, \ldots, n$. Let G_n^* be the empirical distribution function of the X_i^*, $i = 1, \ldots, n$, and let $\alpha_n^{G_n}(t) = \sqrt{n}\big(G_n^*(t) - G_n(t)\big)$ be the bootstrap empirical process. Since $\alpha_n^{G_n}(t) = \alpha_n^*\big(G_n(t)\big)$, it follows from (21) that

$$\|\alpha_n^{G_n} - B_n^*(G_n)\|_{\mathbb{R}} = O_{P^*}\big(\log(n)/\sqrt{n}\big) \quad a.s.. \tag{22}$$

Since Brownian bridge is Hölder continuous a.s. with every exponent $0 < \alpha < 1/2$, we have

$$\big|B_n^*(G_n(x)) - B_n^*(G(x))\big| \leq C_{\alpha,n}|G_n(x) - G(x)|^\alpha \quad a.s., \quad x \in \mathbb{R}, \tag{23}$$

for a sequence of random variables $C_{\alpha,n}$ the conditional distribution of which does not depend on n. Furthermore from Csörgö and Révész (1981, p. 156),

$$\|G_n - G\|_{\mathbb{R}} = O\big((\log\log^+(n)/n)^{1/2}\big) \quad a.s. \tag{24}$$

Using (22), (23) and (24) we compute that for fixed $0 < \alpha < 1/2$,

$$\begin{aligned}
\|\alpha_n^{G_n} - B_n^*(G)\|_{\mathbb{R}} &= \|\alpha_n^*(G_n) - B_n^*(G)\|_{\mathbb{R}} \\
&\leq \|\alpha_n^* - B_n^*\|_{[0,1]} + \|B_n^*(G_n) - B_n^*(G)\|_{\mathbb{R}} \\
&\leq O_{P^*}\big(\log n/\sqrt{n}\big) + |C_{\alpha,n}|\|G_n - G\|_{\mathbb{R}}^{\alpha} \\
&\leq O_{P^*}\big(\log n/\sqrt{n}\big) + O_{P^*}(1)O\big((\log\log^+ n/n)^{\alpha/2}\big) \\
&= O_{P^*}\big(\log(n)/n^{\alpha/2}\big) \quad a.s. \qquad (25)
\end{aligned}$$

Now consider the process

$$Y_{n,0}^{*(j)}(t) = \frac{h^{\beta-3/2}}{g(t)^{1/2}} \int_{\mathbb{R}} K_n^{(j+1)}\Big(\frac{t-x}{h}\Big)\, B_n^*(G(x))\, dx, \quad t \in [0,1].$$

Using (25), the following lemma is proved with the same arguments as Lemma 4 and Corollary 2.

Lemma 7. *Under the assumptions of Theorem 3, we have that*

$$\|Y_n^{*(j)} - Y_{n,0}^{*(j)}\|_{[0,1]} = o_P(c_n^{-1}) \quad a.s.$$

Lemma 7 implies that

$$d\Big(\mathcal{L}^*\big(c_n\|Y_n^{*(j)}\|_{[0,1]}\big), \mathcal{L}^*\big(c_n\|Y_{n,0}^{*(j)}\|_{[0,1]}\big)\Big) \to 0 \quad a.s.,$$

and Lemma 4 that

$$d\Big(\mathcal{L}\big(c_n\|Y_n^{(j)}\|_{[0,1]}\big), \mathcal{L}\big(c_n\|Y_{n,0}^{(j)}\|_{[0,1]}\big) \to 0.$$

But $\mathcal{L}\big(c_n\|Y_{n,0}^{(j)}\|_{[0,1]}\big)$ and $\mathcal{L}^*\big(c_n\|Y_{n,0}^{*(j)}\|_{[0,1]}\big)\Big)$ coincide a.s. An application of the triangle inequality finishes the proof of (18). The final conclusion follows directly from Theorem 1. □

References

Bickel, P. J. and Rosenblatt, M. (1973) On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**, 1071–1095.

Butucea, C. (2004) Deconvolution of Supersmooth Densities with Smooth Noise. *Can. J. Stat.*, **32**, 181–192.

Chaudhuri, P. and Marron, J. S. (2000) Scale space view of curve estimation. *Ann. Statist.*, **28**, 408–428.

Chen, Y. Q., Nissen, P. E., Zhao, G., Zhang, H. W. and Benoni, T. (2000) Chemical composition of

90 F and G disk dwarfs. *Astron. Astrophys. Suppl.*, **141**, 491–506.

Claeskens, G. and van Keilegom, I. (2003) Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist*, **31**, 1852–1884.

Csörgö, M. and Révész, P. (1981) *Strong Approximations in Probability and Statistics.* New York: Academic Press.

Delaigle, A. and Gijbels, I. (2001) Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Stat. Math.*, **56**, 19–47.

Delaigle, A. and Gijbels, I. (2002) Estimation of integrated squared density derivatives from a contaminated sample *J. R. Statist. Soc. B*, **64**, 869–886.

Delaigle, A. and Gijbels, I. (2004) Practical bandwidth selection in deconvolution kernel density estimation *Comput. Statist. Data Anal.*, **45**, 249–267.

Diggle, P. and Hall, P. (1993) A Fourier Approach to nonparametric deconvolution of a density estimate. *J. R. Statist. Soc. Ser. B*, **55**, 523–531.

Edvardsson, B., Andersen, J., Gustafsson, B., Lambert, D. L., Nissen, P. E. and Tomkin, J. (1993) The Chemical Evolution of the Galactic Disk - Part One - Analysis and Results. *Astron. Astroph.*, **275**, 101–152.

Eubank, R. L. and Speckman, P. L. (1993) Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.*, **88**, 1287–1301.

Fan, J. (1991a) On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, **19**, 1257–1272.

Fan, J. (1991b) Asymptotic normality for deconvolution kernel density estimators. *Sankhya Ser. A*, **53**, 97–110.

Fan, Y. and Liu, Y. (1997) A note on asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A*, **59**, 138–141.

Giné, E., Koltchinskii, V. and Sakhanenko, L. (2004) Kernel density estimators: convergence in distribution for weighted sup-norms. *Probab. Theory Relat. Fields*, **130**, 167–198.

Hall, P. (1993) On Edgeworth expansion and bootstrap confidence bands in nonparametric regression. *J. R. Statist. Soc. B*, **55**, 291–304.

Hall, P. and Owen, A. B. (1993) Empirical likelihood confidence bands in density estimation. *J. Com-*

put. Graph. Statist., **2**, 273–289.

Härdle, W. (1989) Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.*, **29**, 163–179.

Hesse, C. H. (1999) Data-driven deconvolution. *J. Nonparametr. Stat.*, **10**, 343–373.

Hesse, C. H. and Meister, A. (2004) Optimal iterative density deconvolution. *J. Nonparametr. Stat.*, **16**, 879–900.

Jørgensen, B. R. (2000) The G dwarf problem. *Astron. Astroph.*, **363**, 947–957.

Karatzas, I. and Shreve, S. E. (1991) *Brownian Motion and Stochastic Calculus*, 2nd edn. New York: Springer.

Komlós, J., Major, P. and Tusnády, G. (1975) An approximation of partial sums of independent random variables and the sample distribution function. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **32**, 111–131.

Lynden-Bell, D. (1975) The chemical evolution of galaxies. *Vistas in Astronomy*, **19**, 299–316.

Mair, B. A. and Ruymgaart, F. H. (1996) Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, **56**, 1424–1444.

Meister, A. (2004) On the effect of misspecifying the error density in a deconvolution problem. *Can. J. Stat.*, **32**, 439–449.

Neumann, M. H. and Polzehl, J. (1998) Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparametr. Statist.*, **9**, 307–333.

Nordström, B., Mayor, M., Andersen, J., Holmberg, J., Pont, F., Jørgensen, B. R., Olsen, E. H., Udry, S. and Mowlavi, N. (2004) The Geneva-Copenhagen survey of the Solar neighbourhood. Ages, metallicities, and kinematic properties of $\sim$ 14000 F and G dwarfs. *Astron. Astroph.*, **418**, 989–1019.

Øksendal, B. (1998), *Stochastic Differential Equations*, 5th edn. Berlin: Springer.

Pagel, B. E. J. (1997) *Nucleosynthesis and Chemical Evolution of Galaxies*. Cambridge (UK): Cambridge University Press.

Pagel, B. E. J. and Tautvaisiene, G. (1995) Chemical evolution of primary elements in the Galactic disc: an analytical model. *Mon. Not. Roy. Astron. Soc.*, **276**, 505–514.

Pensky, M. and Vidakovic, B. (1999) Adaptive wavelet estimator for nonparametric density deconvo-

lution. *Ann. Statist.*, **27**, 2033–2053.

Piterbarg, V. I. and Penskaya, M. Y. (1993) On asymptotic distribution of integrated squared error of an estimate of a component of a convolution. *Math. Methods Statist.*, **2**, 147–164.

Politis, D. N. and Romano, J. P. (1999) Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivariate Anal.*, **68**, 1–25.

Schmidt, M. (1963) The Rate of Star Formation. II. The Rate of Formation of Stars of Different Mass. *Astroph. Journ.*, **137**, 758–769.

Schuster, W. J. and Nissen, P. E. (1989) Uvby-beta photometry of high-velocity and metal-poor stars. II - Intrinsic color and metallicity calibrations. *Astron. Astroph.*, **221**, 65–77.

Shorack, G. R. (1982) Bootstrapping robust regression. *Comm. Statist. A—Theory Methods*, **11**, 961–972.

Sommer-Larsen, J. (1991) On the G-dwarf abundance distribution in the solar cylinder. *Mon. Not. Roy. Astron. Soc.*, **249**, 368–373.

Stefanski, L. and Carroll, R. J. (1990) Deconvoluting kernel density estimators. *Statistics*, **21**, 169–184.

Stute, W. (1982) A law of the logarithm for kernel density estimators. *Ann. Statist.*, **10**, 414–422.

van den Bergh, S. (1962) The frequency of stars with different metal abundances. *Astron. Journ.*, **67**, 486–490.

van Es, B. and Uh, H.-W. (2005) Asymptotic normality of kernel-type deconvolution estimators. *Scand. J. Statist.*, **32**, 467–483.

Xia, Y. (1998) Bias-corrected confidence bands in nonparametric regression. *J. R. Statist Soc. B*, **60**, 797–811